Circular inference predicts nonuniform overactivation and dysconnectivity in brain-wide connectomes

Vincent Bouttier, Suhrit Duttagupta, Sophie Denève, Renaud Jardri

Version of Record: https://www.sciencedirect.com/science/article/pii/S0920996421000311
Manuscript_00f1fded63887050bb0596644870fe17

CIRCULAR INFERENCE PREDICTS NONUNIFORM OVERACTIVATIONS AND DYSCONNECTIVITY IN BRAIN-WIDE CONNECTOMES

Vincent Bouttier [1,2], Suhrit Duttagupta [2], Sophie Denève [2], Renaud Jardri [1,2]

1. Univ Lille, INSERM U1172, CHU Lille, Lille Neurosciences & Cognition Centre (LiNC), Plasticity & SubjectivitY team, 59037 Lille, France.
2. Laboratoire de Neurosciences Cognitives & Computationnelles (LNC²), ENS, INSERM U960, PSL Research University, 75005 Paris, France.

Corresponding authors:
vincent.bouttier@ens.fr; renaud.jardri@chru-lille.fr

ABSTRACT

Schizophrenia is a severe mental disorder whose neural basis remains difficult to ascertain. Among the available pathophysiological theories, recent work has pointed towards subtle perturbations in the excitation-inhibition (E/I) balance within different neural circuits. Computational approaches have suggested interesting mechanisms that can account for both E/I imbalances and psychotic symptoms. Based on hierarchical neural networks propagating information through a message-passing algorithm, it was hypothesized that changes in the E/I ratio could cause a "circular belief propagation" in which bottom-up and top-down information reverberate. This circular inference (CI) was proposed to account for the clinical features of schizophrenia. Under this assumption, this paper examined the impact of CI on network dynamics in light of brain imaging findings related to psychosis. Using brain-inspired graphical models, we show that CI causes overconfidence and overactivation most specifically at the level of connector hubs (e.g., nodes with many connections allowing integration across networks). By also measuring functional connectivity in these graphs, we provide evidence that CI is able to predict specific changes in modularity known to be associated with schizophrenia. Altogether, these findings suggest that the CI framework may facilitate behavioral and neural research on the multifaceted nature of psychosis.

INTRODUCTION

Cognitive dysfunctions (e.g., impaired attention, working memory, or abstract thinking) and aberrant beliefs and perceptions (e.g., delusions and hallucinations) are prevalent features of schizophrenia. Numerous studies have attempted to decipher the neurobiological bases of these symptoms mostly using brain imaging or pharmacological methods. However, given the complexity of the results, psychosis retains much of its mystery. An essential difficulty is due to the absence of a dominant framework able to relate the widely different levels of analysis available. Computational approaches represent a nascent attempt at bridging these gaps (Adams et al., 2013; Anticevic et al., 2015; Fletcher and Frith, 2009; Krystal et al., 2017; Sterzer et al., 2018). In this paper, we propose a new computational method to relate psychosis with impaired global brain dynamics based on two simple hypotheses. First, the brain is an inference machine (Knill and Pouget, 2004; Lochmann and Deneve, 2011). Second, psychosis is associated with imbalances between excitation (E) and inhibition (I) in local neural circuits (Foss-Feig et al., 2017; Jardri et al., 2016; Lisman, 2012; Sohal and Rubenstein, 2019).

We know that structural and functional brain networks exhibit massive changes in patients with schizophrenia (Brandl et al., 2019). This finding is compatible with the common theory assuming that psychotic disorders result directly from anatomical-functional dysconnections (Friston et al., 2016; Friston, 2020; Murray and Anticevic, 2017; Stephan et al., 2009; Yang et al., 2016) and that small functional dysfunctions can easily spread between linked elements within unimpaired complex networks (Carrera and Tononi, 2014; Fornito et al., 2015; Pantano et al., 1986; Price et al., 2001).

However, the exact mechanisms underlying this breakdown of integration between widely distributed brain areas are poorly understood. These impairments do not seem related to macroscopic lesions and are more likely related to subtle and diffuse deficits at the microscale (e.g., impaired neuromodulation or synaptic plasticity and E/I imbalances). Unfortunately, consensus regarding this topic is lacking.

How should the field of computational psychiatry proceed in the face of such uncertainty? One possible strategy is to directly explore the influence of different candidate mechanisms on brain circuits (e.g., using large-scale modeling of intact and impaired neural networks) to attempt to predict (nontrivial) neural and behavioral effects. Another strategy is to set aside the complexity of the real brain in favor of normative models of belief/behavior formation in humans before searching for signatures of these processes in neural signals. Finally, some recent approaches initially proposed normative models but further proposed (highly simplified) neural mechanistic models that may account for aberrant belief formation (Adams et al., 2013; Jardri and Denève, 2013). Notably, these strategies are unlikely to succeed on their own. A successful model should quantitatively (and qualitatively) account for behavioral and neural data, even when tested outside of its “area of comfort” (e.g., the task it was specifically designed for).

A promising framework to achieve such a goal considers the fact that a major brain function is to build internal predictive representations of its uncertain sensory-motor environment (Doya et al., 2007). Roughly speaking, brain circuits would mirror an underlying hierarchy of causes with sensory inputs at the bottom and more abstract knowledge/context at the top (**Fig 1a**). Inference in such a system occurs by integrating information propagated in opposite directions within neural circuits with sensory information “climbing” the hierarchy through feedforward connections, while prior knowledge descends the hierarchy using feedback connections (**Fig 1b**, see also “Summary of methods”). In their simplest expression, these models apply the Bayes theorem in which priors (top-down predictions) and likelihoods

(bottom-up sensory information) are combined with weights corresponding to their reliability. This approach is equivalent to correcting the prior with a prediction error (see also Aitchison and Lengyel, 2017 for a critical discussion regarding Bayesian inference and predictive coding).

While the specific neural mechanisms underlying inference are still highly controversial, the different corresponding computational models have much in common. The brain structure is assumed to represent an underlying probabilistic structure (Parr and Friston, 2018). Neural activity represents probabilities or probabilistic updates (Pouget et al., 2013). Connection strength represents how reliably variables are able to predict each other's state, and inference is performed by propagating local messages (beliefs, predictions, prediction errors, etc.) through these connections. For example, given some sensory evidence for the color green and a prior belief of walking under trees, the up and down propagation of messages allows computing the probability of perceiving leaves in the environment (**Fig 1a,b**).

However, these computational models also differ in the assumed impairments at the roots of aberrant beliefs, such as those that may occur during psychosis. For instance, certain connection types could be disproportionately strong (e.g., an overweighting of top-down messages would result in priors dominating the percept (see Corlett et al., 2018 for a review), which corresponds to changes in the generative model (Parr et al., 2019). Alternatively, we hypothesize that the generative model is unchanged and that the inference mechanism (message-passing scheme) is dysfunctional as follows: messages could be uncontrollably reverberated and amplified through feedforward/feedback loops (**Fig 1c**) and, in turn, drive the perceptual content. Indeed, we previously showed that such a form of *circular inference* (CI) could be a direct consequence of impaired inhibitory control in hierarchical brain circuits (Denève and Jardri, 2016; Jardri and Denève, 2013; Leptourgos et al., 2017).

If valid, such theoretical models should be able to capture individual behavior using a minimal set of parameters. For instance, we found that different levels of CI could account for nonpathological (e.g., illusions (Notredame et al., 2014), bistable perceptions (Leptourgos et al., 2020b, 2020a)) and pathological behaviors, such as the heterogeneous features/dimensions of schizophrenia (Jardri et al., 2017). Fewer studies validating probabilistic models at the neural level have been performed e.g. Powers et al., 2017. A major difficulty is that large-scale brain networks are far more complex than the simple hierarchical chains used in toy examples or to describe experimentally designed tasks.

The goal of this paper is to provide a proof of concept for extending the CI model to brain-wide neural activity with possible applications in the context of psychosis. More precisely, we show that it is possible to apply CI to simplified brain-like (abstract) graphs or brain-based connectomes (Bullmore and Bassett, 2011). We predicted the basic impairments due to CI in these graphs at both the activity level and the functional connectivity level. Finally, we compared these predictions with common fMRI findings of dysconnectivity, as observed in schizophrenia.

INSERT FIGURE 1 ABOUT HERE

SUMMARY OF METHODS

This section succinctly describes how the graphs were generated and how their activity was stimulated. For more details, see **Supplementary Material**. The code (in Python) is available online at github.com/VincentBt/.

We randomly generated modular small-world graphs (two common properties of brain-like networks - **Fig 2a,b,d**), which we call "abstract graphs". Nodes within the graphs were assumed to receive randomly fluctuating, temporally smooth inputs (insets in **Fig 2c**). The goal of using randomly generated graphs was to predict dynamic properties independent of the specific structure of the network. We used random input patterns to mimic resting-state brain activity as opposed to task-based functional patterns.

Inference was performed in these graphs by continuously propagating messages in multiple directions along the links using a local message-passing algorithm called *belief propagation* (BP) (Bishop, 2006; Friston et al., 2017). The confluence of messages in a given node was used to compute its belief, to be understood as a local estimate of the probability that the binary variable encoded by the node is 1 ($b = p(X=1)$), given the currently available evidence. Thus, the nodes in the generated graph constantly attempt to reach an agreement by exchanging predictions regarding each other's states. The "sensory" evidence provided to the network (the random inputs) smoothly changed over time, as did the beliefs, as exemplified in **Fig 2c**. Importantly, the reverberation of messages is avoided in BP by removing the message previously sent in the opposite direction from each message, which is carried out by inhibition in our proposed neural implementation (**Fig 1b**). For example, when a tree predicts leaves, leaves should not subsequently predict a tree by total circular reasoning (this circularity is illustrated in **Fig 1c**). To implement CI with increasing severity, we progressively decreased this correction using parameter α representing the level of inhibitory control and using values from 60% (strong circular reasoning - impaired inhibitory control) to 100% (BP - perfect inhibitory control).

INSERT FIGURE 2 ABOUT HERE

We analyzed the statistics of the beliefs generated in these graphs for normal inference (BP) and increasing amounts of pathological inference (CI). As a sanity check, the predictions obtained by investigating the abstract graphs were reproduced using a specific but more realistic brain-based graph or "realistic connectome" by utilizing the set of reconstructed group-averaged fiber tracts from the open-access HCP-842 MRI atlas (Yeh et al., 2018) combined with a collection of 86 anatomical parcels (see **Supplementary Table**) taken from the AAL2 atlas (Rolls et al., 2015). We referred to a canonical division of nodes in a priori communities to define structural modules (also called clusters, communities, or groups) as proposed by Bertolero et al., 2018 (see **Fig S1**).

Nodes are divided into the following three categories: connector hubs, local hubs, and

other nodes. Connector hubs are nodes with connections that are diversely distributed across modules. Local hubs are nodes that are highly connected within their own module (and are not connector hubs). Other nodes are all nodes that are not connector hubs or local hubs. Connector hubs are defined based on the *participation coefficient* (a measure of the diversity of intermodular connections of the node), and local hubs are defined based on the *within-community strength* (a measure of the locality of the node through its intramodular connections). See **Supplementary Material** for a formal definition of the node types and graph metrics.

NEURAL INTERPRETATION

To interpret the graph dynamics in neural terms, we need to decide how beliefs translate into neural activity, which is a topic that is still controversial. For simplicity, it was assumed that activity in a brain parcel covaried with the confidence level of the corresponding node (i.e., the absolute value of $\log(b/(1-b))$ where belief b is the probability that the binary node is in state 1 - see **Supplementary Material** for more detail). Thus, the more certain a node was of its variable state at a given time (in the case of binary variables, the closer its belief was to 0 or 1), the more active the node was. Note that some studies identified a link between neural activity and surprise (Schwartenbeck et al., 2016), which approximately corresponds to large temporal fluctuations in beliefs. In our simulations, the two quantities were strongly correlated, and both predicted essentially the same effects on functional connectivity (see also **Supplementary Material**). Assuming a neural representation of surprise instead of beliefs would not change any of the conclusions presented here.

As previously mentioned, one cannot make a direct and naive parallel between BP (or CI) and the dynamics of a brain-like network with a matching connectivity structure. If the graph links are indeed analogous to recurrent connections between corresponding neural populations, reciprocal anatomical connections cause positive feedback, i.e., reverberation of messages. The cancellation of the reverberated part of the messages in BP is proposed to be carried out by inhibitory control, which can occur locally as shown in **Fig 1b** or through long-range connections (Leptourgos et al., 2017). This would imply that BP corresponds to the dynamics of a superbalanced brain in which recurrent loops are constantly controlled by tight local inhibition (Denève and Machens, 2016). Presumably, relaxing this inhibitory control results in an increased CI, which is measurable at the behavioral level, even if signatures of this process at the neural level remain to be found.

CIRCULAR INFERENCE IN ABSTRACT GRAPHS

We first report the effects observed on randomly generated graphs. We observed that CI induces overconfidence and, thus, generates an excess of neural activity. On average, the CI-generated confidence levels are indeed higher as reflected by a sigmoidal relationship between the CI and BP-computed posterior probabilities (**Fig 3a, upper panel**). Similarly, the distribution of beliefs among all nodes extends further towards extreme values at higher levels of CI (**Fig 3b, upper panel**), and this result persists when bounding the belief under belief

propagation between 0.4 and 0.6 (**Fig S2**). Thus, while BP generates graded beliefs in proportion with the weak and/or contradictory evidence provided to the network (i.e., fluctuating inputs), CI causes more extreme levels of certainty.

In reality, this average relationship at the network scale hides a large amount of heterogeneity (**one-way ANOVA**, **F(2, 894) = 215, p<0.001**). The effect in some nodes is much stronger than that in other nodes (dependent on the local structure of the network as described later in **Fig 4b**). In the most affected nodes, CI causes beliefs to saturate to extreme values in a large portion of what should normally be their response range (**Fig 3c, upper panel**). Thus, these nodes not only are aberrantly confident but also become insensitive to small fluctuations in their input messages and, thus, are presumably unable to transfer information to nodes downstream in the network. This finding suggests that CI not only causes overconfidence but also, somewhat counterintuitively, weakens the communication between nodes.

INSERT FIGURE 3 ABOUT HERE

Upon closer examination, one finds that the variations in overconfidence induced by CI are explained by only a few properties characterizing the centrality of a node within the graph (**Fig S3**). The nodes most affected by CIs are *connector hubs* whose connections are diversely distributed across modules (**post hoc comparisons** using t-tests for independent samples revealed that connector hubs exhibited significantly higher confidence than local hubs, $\mathbf{p = 2.58\ e^{-22}}$, or other nodes, $\mathbf{p = 1.69\ e^{-88}}$; it should be noted that local hubs, which are nodes that are highly connected within their own module, also significantly differ from the other nodes, $\mathbf{p = 8.6\ e^{-20}}$; see also **Fig 4**). For a formal definition of connector hubs, local hubs, and other nodes, see **Supplementary Material**. These results concerning overconfidence also apply to overactivation (excess of neural activity). Indeed, overconfidence and overactivation have the same definition in the model (see **Supplementary Material**). Consequently, there is an overall overactivation of the network, especially in the network hubs.

INSERT FIGURE 4 ABOUT HERE

Since connector hubs exert maximal control over long-range communication within and between modules, one could expect that as the severity of CI increases, the network becomes more strongly modular with weaker functional interactions at long-range relative to short-range. These expected consequences of CI are confirmed when directly measuring functional connectivity based on the graph responses (**Fig 5a**). Here, functional connectivity is defined as the amount of correlation between all pairs of nodes (see **Supplementary Material**). Measuring such functional connectivity provides an approximate idea of the underlying structure (anatomical connectivity) of the network. As expected, truly connected nodes exhibit strong correlations (increased by CI, see **Fig S4**), while nodes separated by

longer paths exhibit lower levels of correlation.

Finally, as predicted from the effects of CI on hubs, the degree of modularity of the network (defined as the intra-inter modular ratio of the functional connectome, see **Supplementary Material**) significantly increases with CI as functional connectivity increases within a given module but comparatively decreases between different modules (**Fig 5b**). Thus, the network becomes less able to process information at the global scale while comparatively sparing intramodular (local) communication.

INSERT FIGURE 5 ABOUT HERE

These predictions might appear counterintuitive given that reverberation could appear to increase rather than decrease the amount of global communication between nodes. However, these simulation results show that the crucial element in transferring information between different parts of the graphs is to keep beliefs graded and driven by external inputs and afferent messages rather than saturated by internal recurrent dynamics. Thus, CI paradoxically predicts both overconfidence (especially in variables represented in associative or multimodal brain areas) and a deficit in global processing of information with a marked decrease in long-range (functional) connectivity relative to local connectivity.

CIRCULAR INFERENCE IN A MORE REALISTIC BRAIN CONNECTOME

As observed in the randomly generated graphs, the implementation of CI in a more realistic connectome resulted in overconfidence, especially in the connector hubs (**Figs 3 & 4 lower panels**, **$F(2, 930) = 903$, $p<0.001$**; post hoc comparisons: connector hubs > local hubs: **$p=4.0\ e^{-12}$**; connector hubs > other nodes: **$p=6.0\ e^{-323}$**: local hubs > other nodes: **$p= 3.5\ e^{-186}$**). A linear model was trained to predict overactivation induced by circular inference (see **Supplementary Material** for a formal definition of overactivation) based on various graph measures characterizing the centrality of each node (which is a proxy for the amount of control a node has over communication in the network). This linear model trained on randomly generated graphs or “abstract graphs” accurately predicted the level of overactivation in each node in the connectome (up to a constant of proportionality, **Pearson’s *r* correlation = 0.97, $p= 1.8\ e^{-52}$**; see also **Fig S5**).

We ranked the 86 parcels of this connectome according to the level of overactivation caused by CI. Represented on the same scale, we can observe how strongly the *degree centrality and, to* a lesser extent, the *participation coefficient* and the *within-community strength* are correlated with overactivation (**Fig S6** - see **Supplementary Material** for a formal definition of these three graph metrics).

Finally, the results showing functional connectivity impairments in the abstract graphs were also observed in the real connectome as circular inference increases the modularity in the functional network (**Fig 5 lower panels**), i.e., the long-range connections between modules were more strongly affected than the short-range connections within a module.

DISCUSSION AND PERSPECTIVES

In this paper, we explored how disruptions of E/I homeostasis in a graph model of the brain can change the dynamics of the network and modify information propagation, eventually leading to psychotic-like symptoms. We developed a whole-brain computational model relating the probabilities encoded by a population to their neural activity using efficient coding principles. This normative approach allowed us to simulate the network dynamics and reproduce some results from the connectomics literature concerning psychosis, specifically the generation of overconfidence and overactivations (centered on hubs) and the inability to maintain an efficient modular small-world architecture (i.e., the increase in modularity in the functional connectome).

Breakdowns in the E/I balance at the microcircuit level are considered major alterations in neurodevelopmental disorders (Foss-Feig et al., 2017; Sohal and Rubenstein, 2019). Notably, an increased E/I ratio was not only proposed to drive psychotic features in schizophrenia but could also be involved in more acute disorders such as anti-$NMDA_R$ encephalitis (Parenti et al., 2016). The CI framework, which is based on impaired inhibitory control, appears particularly useful for modeling psychosis across diagnosis categories and has already received some behavioral support. For instance, the overconfidence due to CI is compatible with prior theoretical results (Jardri and Denève, 2013) and the model fitting of the *Jumping-to-conclusions* reasoning bias in schizophrenia patients (Jardri et al., 2017), which is usually also correlated with delusional severity (Dudley et al., 2016; Glöckner and Moritz, 2009; Moritz and Woodward, 2005).

The present study extends this literature by showing that CI applied in a brain-like network can generate a nonuniform distribution of aberrantly strong beliefs. More specifically, we showed that belief saturation (excessively high levels of confidence) observed under CI in network hubs prevents the hubs from properly transferring information, which is expressed as intense overactivation. Previous reports have suggested that these hubs play a pivotal role in psychiatric disorders in general (Crossley et al., 2014; Fornito et al., 2015) and schizophrenia in particular (Crossley et al., 2016; van den Heuvel et al., 2013). For instance, based on fMRI symptom-capture studies, it is known that patients with psychosis exhibit specific patterns of hyperactivation during hallucinatory experiences (Ćurčić-Blake et al., 2017; Jardri et al., 2011; Sommer et al., 2008). These signal changes are localized in not only essential hubs that constitute a part of the speech-related network when hallucinations occur in the verbal domain (mainly in the inferior frontal gyrus and the temporoparietal junction) but also amodal epicenters involved in contextual memory, such as the hippocampal complex.

Thus, specific dysfunctions in connector hubs appear compatible with the clinical richness and cross-domain impairments of schizophrenia. This nonuniform distribution of beliefs/activations may account for the apparently unrelated observations that psychosis, on the one hand, can generate unshakable beliefs but, on the other hand, may result in impaired information processing.

As a consequence of such localized impairments in hubs, we observe a shift from global to local connectivity (increased short- relative to long-range functional connectivity) and widely distributed miscommunication in the network, which also appears compatible with previous reports investigating schizophrenia (Li et al., 2017; Xiang et al., 2019; Zalesky et al., 2011). These changes in the brain network topology were previously found to be linked with

psychotic symptoms. For instance, functional hyperconnectivity was observed between different parts of the language network (module AUD in **Fig S1**) in patients suffering from auditory-verbal hallucinations, while hypoconnectivity was found with other distant brain areas (Shinn et al., 2013). Interestingly, these functional dysconnectivity patterns were found to be state-dependent and modulated by antipsychotic medication (Hadley et al., 2016).

The modular topology is known to physiologically vary across an individual's lifespan and notably optimize and correlate with cognitive efficiency during adolescence (Baum et al., 2017). Proper functioning of the brain (i.e., effective segregation and integration of information processing) depends on these short- and long-range connections. Interestingly, synaptic elimination and late E/I balance adjustments contribute to adolescent brain maturation (Selemon, 2013), which is considered to be a critical developmental period for schizophrenia onset (Paus et al., 2008; Rolls and Deco, 2011). Precisely, the transition to psychosis has been shown to be associated with a preferential reduction in long-range connections in patients compared with nonclinical relatives, who shared a genetic vulnerability to schizophrenia but did not develop the disorder (Guo et al., 2014).

It is usually well accepted that global changes in the structural connectivity of the brain represent a pathological hallmark of several neuropsychiatric disorders (Lord et al., 2017). However, our results also suggest that the inability to properly integrate information in different brain areas could be partially due to pure impaired dynamics (i.e., even without modifications of the structural connections), representing a particularly interesting process accounting for acute psychotic manifestations as such manifestations can be observed beyond the schizophrenia spectrum. These predictions of the CI model (in which we alter the inference mechanism but not the anatomical graph) are notably compatible with data from animals exposed to ketamine (an NMDA antagonist - (Voss et al., 2012)) or E/I changes following chemogenetic manipulation (Markicevic et al., 2020) in which a reduction in long-range connectivity was observed. These predictions are also consistent with recent brain imaging findings in patients suffering from anti-$NMDA_R$ encephalitis who exhibit multiple focal increases in neural activity (Miao et al., 2020) and a significant decrease in the strength of long-range connections (Peer et al., 2017), even though the structure of the anatomical graph is preserved. Of course, we cannot exclude the possibility that persistent functional changes may lead to plastic structural impairments in the long run if not properly fixed, which could be a nice development of the model.

We would like to acknowledge the very preliminary nature of the present work. We intended to provide a proof of concept rather than a detailed framework of spontaneous brain dynamics based on the connectome as we use random connections, random inputs, etc. For instance, spontaneous activity is triggered in the graphs by using random inputs injected in all nodes. In reality, even spontaneous activity is likely to have anatomically and spatially more limited sources (Uddin, 2020). Better identification of these sources in the future might explain more specific patterns in brain activity, such as those in the default-mode network (Fox et al., 2005), and improve the performance of the model in predicting which brain areas should be overactivated. Furthermore, more detailed information could be obtained by using the weights of the connectome. Finally, the current model is restricted to binary variables; subsequent work could adapt the circular inference framework to continuous variables (see **Supplementary Material**).

In addition, several issues remain unanswered and may benefit from further

clarification. Among these possible tracks, we should mention the effect of implementing different types of inference loops in the graph (e.g., ascending/descending), more precise exploration of the neural hierarchy, the trigger of specific subnetworks (e.g., thalamocortical loops or the hippocampal-prefrontal pathway) or even fitting fMRI data from patients with various psychotic symptoms, e.g., suffering from schizophrenia or anti-$NMDA_R$ encephalitis.

Nevertheless, the current findings suggest that the same parametric model (CI) could fit behavioral (Jardri et al., 2017) and neural (this study) data and, thus, pave the way for transdiagnostically linking neural signals with psychosis.

FIGURE CAPTIONS

Figure 1: Principles of belief propagation and circular inference. (A) Toy example of a hierarchical causal model of three nodes representing hidden variables. The sensory input for the color green is given at the bottom of the hierarchy, while the prior expectation of a tree is given at the top. The beliefs in each node are shown in green. **(B)** A possible implementation of the belief propagation algorithm by a neural network. The information shared between the different nodes of the network is under the control of inhibitory interneurons (shown in red), which remove redundant information from messages. **(C)** In the case of circular inference, an impairment in the interneurons (dotted lines) causes an uncontrolled reverberation of messages in the network, leading to aberrant beliefs (depicted here with green halos).

Figure 2: Running belief propagation in abstract small-world networks. Graphical networks are randomly generated with small-world properties and a modular structure consisting of 4 modules with 8 nodes per module. **(A)** Adjacency matrix of one of the networks generated. **(B)** Graphical representation of the network. Each of the 4 modules gathers nodes of a given color. All nodes receive a randomly fluctuating, temporally smooth input (insets). **(C)** Illustration of the temporal evolution of beliefs (probability estimates) in the network using proper inference (i.e., belief propagation). We present the beliefs in 3 nodes randomly selected from the graph. **(D)** Graphical representation of the same network presented in (B) but using a yellow-to-red color code to reflect the participation coefficient and node size for the degree.

Figure 3: Effect of circular inference on beliefs in the abstract small-world networks and the real connectome. The results based on belief propagation in randomly generated small-world modular graphs are presented in the upper panels, and those for the realistic connectome network are presented in the lower panels. **(A)** Plot of the posterior probabilities as measured by circular inference (CI) against the same probabilities from belief propagation (BP), averaged. Decreasing the level of inhibitory control (i.e., increasing the level of CI) causes the nodes to have greater confidence compared to BP (where = 100% represents BP). **(B)** Distribution of beliefs in a single node while varying the degree of circularity. Lower inhibitory control causes more extreme beliefs. **(C)** A comparison of the beliefs under CI in one node against the same beliefs under BP. CI causes the nodes to saturate towards more extreme beliefs.

Figure 4: Factors causing overconfidence due to circular inference. (A) The following three types of nodes were considered in the graphs: connector hubs (nodes with connections that are diversely distributed across modules – shown in green); local hubs (nodes that are highly connected within their own module – shown in orange) and other nodes (shown in blue). **(B)** Overconfidence measured in the random graphs and realistic connectome according to the type of nodes (rain-cloud plots), for □ = 60%. Connector hubs are significantly more overconfident than local hubs, which are significantly more overconfident than the other nodes in the network. The results are the same when examining overactivation.

Figure 5: **Functional connectivity under circular inference.** The results for randomly generated small-world graphs are presented in the left panels, and those for the realistic connectome network are presented in the right panels. **(A)** Functional connectivity matrix network as measured by the activation function applied over the beliefs in the network. Regarding the real connectome, the modules are presented in the following order: auditory, sensorimotor, visual, dorsal attention, salience, frontoparietal, default-mode, subcortical, and finally, nodes not attributed to a specific module. **(B)** We explored the ratio between the number of intramodular connections and the number of intermodular connections in the functional network. BP was chosen as the reference to explore the impact of varying the degree of circularity. The ratio significantly increases when we decrease inhibitory control, rendering the network more modular. B left corresponds to A left (abstract graph), and B right corresponds to A right (real connectome).

REFERENCES

Adams, R.A., Stephan, K.E., Brown, H.R., Frith, C.D., Friston, K.J., 2013. The computational anatomy of psychosis. Front Psychiatry 4, 47. https://doi.org/10.3389/fpsyt.2013.00047

Aitchison, L., Lengyel, M., 2017. With or without you: predictive coding and Bayesian inference in the brain. Curr Opin Neurobiol 46, 219–227. https://doi.org/10.1016/j.conb.2017.08.010

Anticevic, A., Murray, J.D., Barch, D.M., 2015. Bridging levels of understanding in schizophrenia through computational modeling. Clin Psychol Sci 3, 433–459. https://doi.org/10.1177/2167702614562041

Baum, G.L., Ciric, R., Roalf, D.R., Betzel, R.F., Moore, T.M., Shinohara, R.T., Kahn, A.E., Vandekar, S.N., Rupert, P.E., Quarmley, M., Cook, P.A., Elliott, M.A., Ruparel, K., Gur, R.E., Gur, R.C., Bassett, D.S., Satterthwaite, T.D., 2017. Modular segregation of structural brain networks supports the development of executive function in youth. Curr Biol 27, 1561-1572.e8. https://doi.org/10.1016/j.cub.2017.04.051

Bertolero, M.A., Yeo, B.T.T., Bassett, D.S., D'Esposito, M., 2018. A mechanistic model of connector hubs, modularity and cognition. Nat Hum Behav 2, 765–777. https://doi.org/10.1038/s41562-018-0420-6

Bishop, C.M., 2006. Pattern Recognition and Machine Learning. Springer.

Boerlin, M., Machens, C.K., Denève, S., 2013. Predictive coding of dynamical variables in balanced spiking networks. PLoS Comput. Biol. 9, e1003258. https://doi.org/10.1371/journal.pcbi.1003258

Brandl, F., Avram, M., Weise, B., Shang, J., Simões, B., Bertram, T., Ayala, D.H., Penzel, N., Gürsel, D.A., Bäuml, J., Wohlschläger, A.M., Vukadinovic, Z., Koutsouleris, N., Leucht, S., Sorg, C., 2019. Specific Substantial Dysconnectivity in Schizophrenia: A Transdiagnostic Multimodal Meta-analysis of Resting-State Functional and Structural Magnetic Resonance Imaging Studies. Biological Psychiatry 85, 573–583. https://doi.org/10.1016/j.biopsych.2018.12.003

Bullmore, E.T., Bassett, D.S., 2011. Brain graphs: graphical models of the human brain connectome. Annu Rev Clin Psychol 7, 113–140. https://doi.org/10.1146/annurev-clinpsy-040510-143934

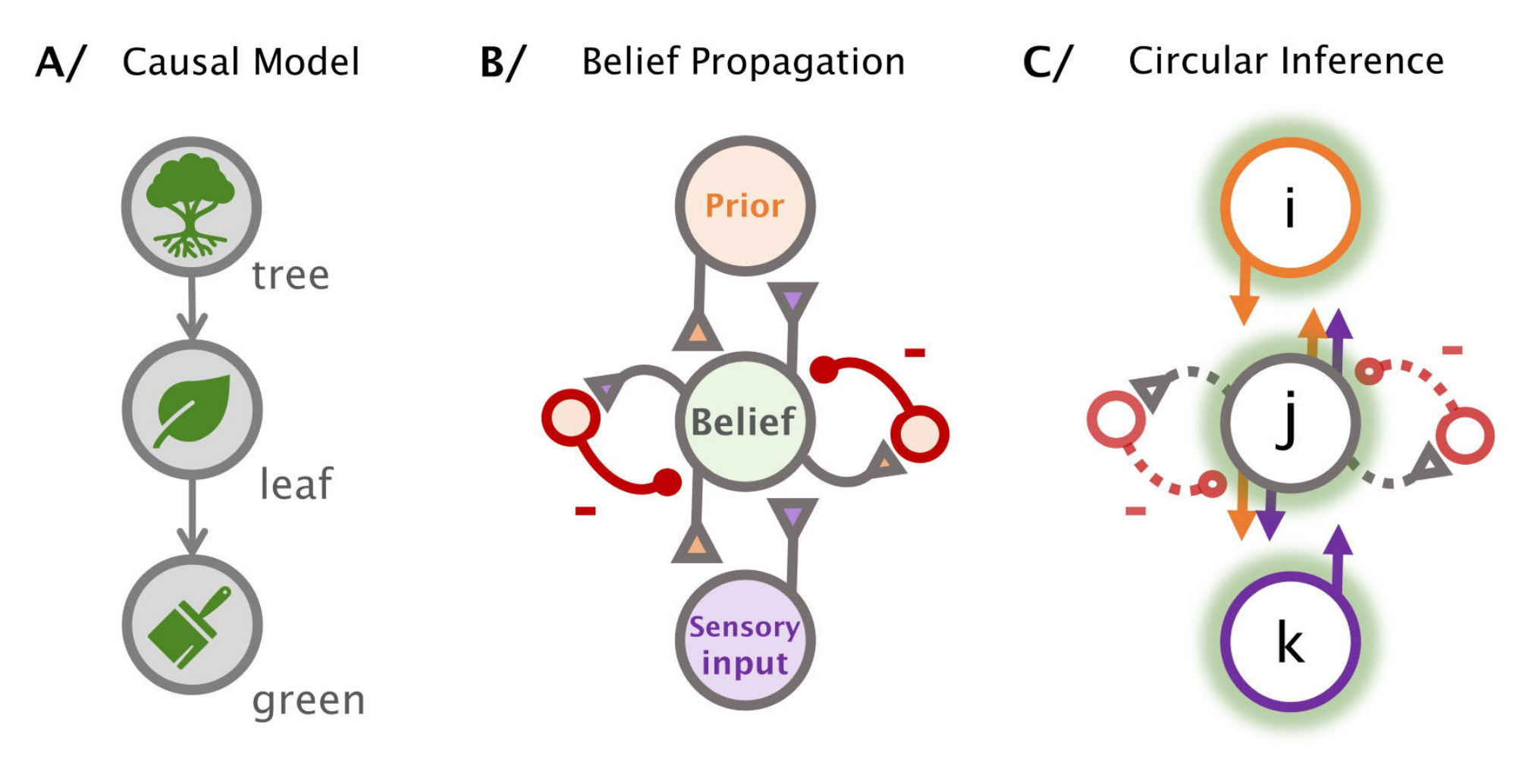
Carrera, E., Tononi, G., 2014. Diaschisis: past, present, future. Brain 137, 2408–2422.

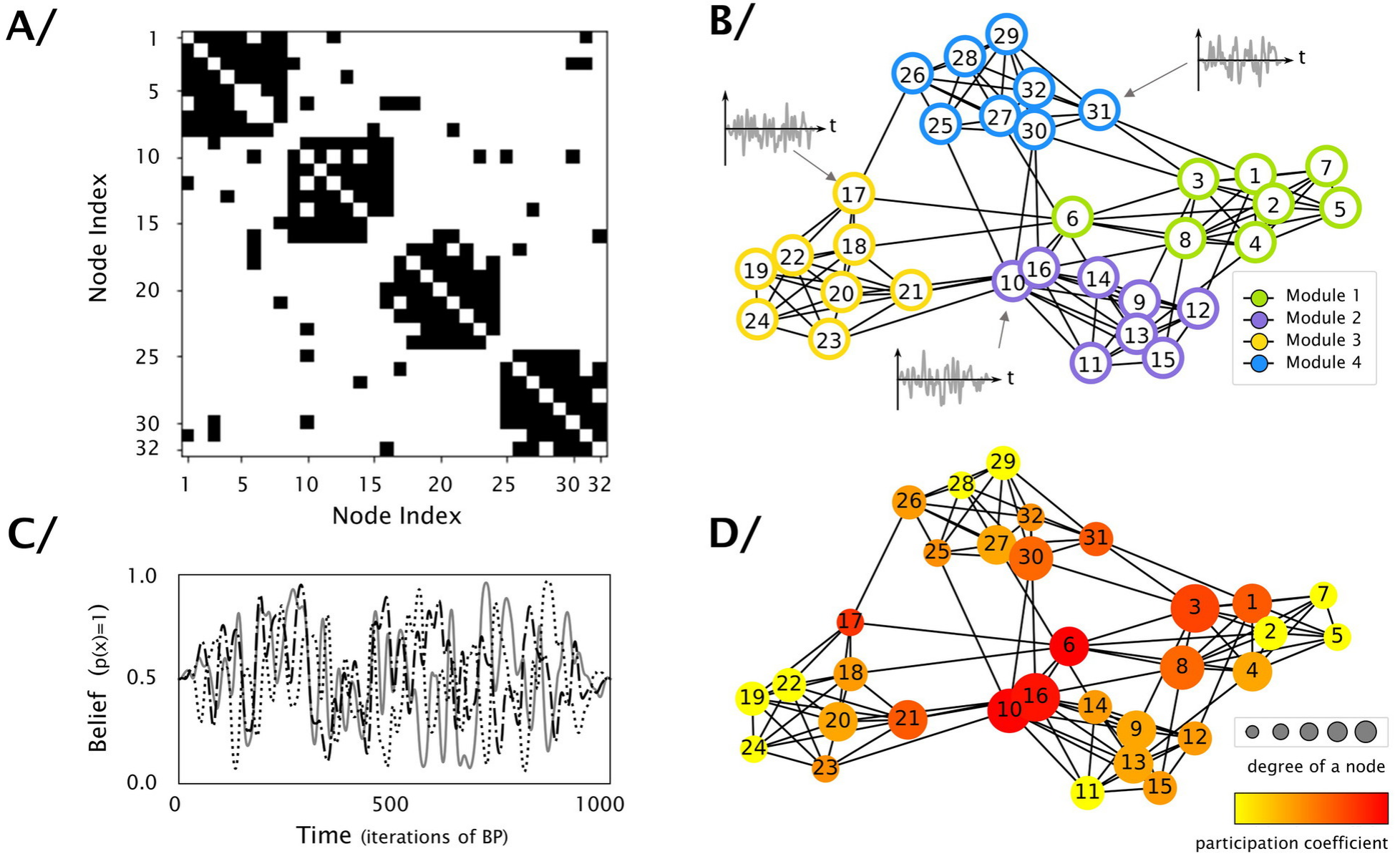
https://doi.org/10.1093/brain/awu101
Corlett, P.R., Horga, G., Fletcher, P.C., Alderson-Day, B., Schmack, K., Powers, A.R., 2018. Hallucinations and strong priors. Trends Cogn Sci (Regul Ed). https://doi.org/10.1016/j.tics.2018.12.001
Crossley, N.A., Mechelli, A., Ginestet, C., Rubinov, M., Bullmore, E.T., McGuire, P., 2016. Altered Hub Functioning and Compensatory Activations in the Connectome: A Meta-Analysis of Functional Neuroimaging Studies in Schizophrenia. Schizophr Bull 42, 434–442. https://doi.org/10.1093/schbul/sbv146
Crossley, N.A., Mechelli, A., Scott, J., Carletti, F., Fox, P.T., McGuire, P., Bullmore, E.T., 2014. The hubs of the human connectome are generally implicated in the anatomy of brain disorders. Brain 137, 2382–2395. https://doi.org/10.1093/brain/awu132
Ćurčić-Blake, B., Ford, J.M., Hubl, D., Orlov, N.D., Sommer, I.E., Waters, F., Allen, P., Jardri, R., Woodruff, P.W., David, O., Mulert, C., Woodward, T.S., Aleman, A., 2017. Interaction of language, auditory and memory brain networks in auditory verbal hallucinations. Prog Neurobiol 148, 1–20. https://doi.org/10.1016/j.pneurobio.2016.11.002
Denève, S., Jardri, R., 2016. Circular inference: mistaken belief, misplaced trust. Curr Opin Behav Sci 11, 40–48. https://doi.org/10.1016/j.cobeha.2016.04.001
Denève, S., Machens, C.K., 2016. Efficient codes and balanced networks. Nat Neurosci 19, 375–382. https://doi.org/10.1038/nn.4243
Doya, K., Ishii, S., Pouget, A., Rao, R.P.N., 2007. Bayesian Brain. The MIT Press, Cambridge, MA.
Dudley, R., Taylor, P., Wickham, S., Hutton, P., 2016. Psychosis, Delusions and the "Jumping to Conclusions" Reasoning Bias: A Systematic Review and Meta-analysis. Schizophr Bull 42, 652–665. https://doi.org/10.1093/schbul/sbv150
Fletcher, P.C., Frith, C.D., 2009. Perceiving is believing: a Bayesian approach to explaining the positive symptoms of schizophrenia. Nat Rev Neurosci 10, 48–58. https://doi.org/10.1038/nrn2536
Fornito, A., Zalesky, A., Breakspear, M., 2015. The connectomics of brain disorders. Nat Rev Neurosci 16, 159–172. https://doi.org/10.1038/nrn3901
Foss-Feig, J.H., Adkinson, B.D., Ji, J.L., Yang, G., Srihari, V.H., McPartland, J.C., Krystal, J.H., Murray, J.D., Anticevic, A., 2017. Searching for Cross-diagnostic Convergence: Neural Mechanisms Governing Excitation and Inhibition Balance in Schizophrenia and Autism Spectrum Disorders. Biol Psychiatry 81, 848–861. https://doi.org/10.1016/j.biopsych.2017.03.005
Fox, M.D., Snyder, A.Z., Vincent, J.L., Corbetta, M., Essen, D.C.V., Raichle, M.E., 2005. The human brain is intrinsically organized into dynamic, anticorrelated functional networks. PNAS 102, 9673–9678. https://doi.org/10.1073/pnas.0504136102
Friston, K., Brown, H.R., Siemerkus, J., Stephan, K.E., 2016. The dysconnection hypothesis (2016). Schizophr Res 176, 83–94. https://doi.org/10.1016/j.schres.2016.07.014
Friston, K.J., 2020. Bayesian Dysconnections. AJP 177, 1110–1112. https://doi.org/10.1176/appi.ajp.2020.20091421
Friston, K.J., Parr, T., de Vries, B., 2017. The graphical brain: Belief propagation and active inference. Netw Neurosci 1, 381–414. https://doi.org/10.1162/NETN_a_00018
Glöckner, A., Moritz, S., 2009. A fine-grained analysis of the jumping-to-conclusions bias in schizophrenia: Data-gathering, response confidence, and information integration. Judgment and Decision Making 4, 587–600.
Guo, S., Palaniyappan, L., Yang, B., Liu, Z., Xue, Z., Feng, J., 2014. Anatomical Distance Affects Functional Connectivity in Patients With Schizophrenia and Their Siblings. Schizophr Bull 40, 449–459. https://doi.org/10.1093/schbul/sbt163
Hadley, J.A., Kraguljac, N.V., White, D.M., Ver Hoef, L., Tabora, J., Lahti, A.C., 2016. Change in brain network topology as a function of treatment response in schizophrenia: a longitudinal resting-state fMRI study using graph theory. npj Schizophrenia 2, 1–7. https://doi.org/10.1038/npjschz.2016.14
Jardri, R., Denève, S., 2013. Circular inferences in schizophrenia. Brain 136, 3227–3241. https://doi.org/10.1093/brain/awt257
Jardri, R., Duverne, S., Litvinova, A.S., Denève, S., 2017. Experimental evidence for circular inference in schizophrenia. Nat Commun 8, 14218. https://doi.org/10.1038/ncomms14218
Jardri, R., Hugdahl, K., Hughes, M., Brunelin, J., Waters, F., Alderson-Day, B., Smailes, D., Sterzer, P., Corlett, P.R., Leptourgos, P., Debbané, M., Cachia, A., Denève, S., 2016. Are hallucinations due to an imbalance between excitatory and inhibitory influences on the brain? Schizophr Bull 42, 1124–1134. https://doi.org/10.1093/schbul/sbw075

Jardri, R., Pouchet, A., Pins, D., Thomas, P., 2011. Cortical activations during auditory verbal hallucinations in schizophrenia: a coordinate-based meta-analysis. Am J Psychiatry 168, 73–81. https://doi.org/10.1176/appi.ajp.2010.09101522
Knill, D.C., Pouget, A., 2004. The Bayesian brain: the role of uncertainty in neural coding and computation. Trends Neurosci 27, 712–719. https://doi.org/10.1016/j.tins.2004.10.007
Krystal, J.H., Murray, J.D., Chekroud, A.M., Corlett, P.R., Yang, G., Wang, X.-J., Anticevic, A., 2017. Computational psychiatry and the challenge of schizophrenia. Schizophr Bull 43, 473–475. https://doi.org/10.1093/schbul/sbx025
Leptourgos, P., Bouttier, V., Jardri, R., Denève, S., 2020a. A functional theory of bistable perception based on dynamical circular inference. PLoS Comput Biol 16, e1008480. https://doi.org/10.1371/journal.pcbi.1008480
Leptourgos, P., Denève, S., Jardri, R., 2017. Can circular inference relate the neuropathological and behavioral aspects of schizophrenia? Curr Opin Neurobiol 46, 154–161. https://doi.org/10.1016/j.conb.2017.08.012
Leptourgos, P., Notredame, C.-E., Eck, M., Jardri, R., Denève, S., 2020b. Circular inference in bistable perception. Journal of Vision 20, 12–12. https://doi.org/10.1167/jov.20.4.12
Li, P., Fan, T.-T., Zhao, R.-J., Han, Y., Shi, L., Sun, H.-Q., Chen, S.-J., Shi, J., Lin, X., Lu, L., 2017. Altered Brain Network Connectivity as a Potential Endophenotype of Schizophrenia. Scientific Reports 7, 5483. https://doi.org/10.1038/s41598-017-05774-3
Lisman, J., 2012. Excitation, inhibition, local oscillations, or large-scale loops: what causes the symptoms of schizophrenia? Curr Opin Neurobiol 22, 537–544. https://doi.org/10.1016/j.conb.2011.10.018
Lochmann, T., Deneve, S., 2011. Neural processing as causal inference. Curr Opin Neurobiol 21, 774–781. https://doi.org/10.1016/j.conb.2011.05.018
Lord, L.-D., Stevner, A.B., Deco, G., Kringelbach, M.L., 2017. Understanding principles of integration and segregation using whole-brain computational connectomics: implications for neuropsychiatric disorders. Philos Transact A Math Phys Eng Sci 375. https://doi.org/10.1098/rsta.2016.0283
Markicevic, M., Fulcher, B.D., Lewis, C., Helmchen, F., Rudin, M., Zerbi, V., Wenderoth, N., 2020. Cortical Excitation:Inhibition Imbalance Causes Abnormal Brain Network Dynamics as Observed in Neurodevelopmental Disorders. Cereb Cortex 30, 4922–4937. https://doi.org/10.1093/cercor/bhaa084
Miao, A., Liu, Q., Li, Z., Liu, W., Wang, L., Ge, J., Yu, C., Wang, Y., Huang, S., Yu, Y., Shi, Q., Sun, J., Wang, X., 2020. Altered cerebral blood flow in patients with anti-NMDAR encephalitis. J. Neurol. 267, 1760–1773. https://doi.org/10.1007/s00415-020-09747-x
Mooij, J.M., Kappen, H.J., 2005. On the properties of the Bethe approximation and loopy belief propagation on binary networks. J. Stat. Mech. 2005, P11012–P11012. https://doi.org/10.1088/1742-5468/2005/11/P11012
Moritz, S., Woodward, T.S., 2005. Jumping to conclusions in delusional and non-delusional schizophrenic patients. Br J Clin Psychol 44, 193–207. https://doi.org/10.1348/014466505X35678
Muldoon, S.F., Bridgeford, E.W., Bassett, D.S., 2016. Small-World Propensity and Weighted Brain Networks. Sci Rep 6, 22057. https://doi.org/10.1038/srep22057
Murphy, K., Weiss, Y., Jordan, M.I., 1999. Loopy Belief Propagation for Approximate Inference: An Empirical Study, in: UAI.
Murray, J.D., Anticevic, A., 2017. Toward understanding thalamocortical dysfunction in schizophrenia through computational models of neural circuit dynamics. Schizophr Res 180, 70–77. https://doi.org/10.1016/j.schres.2016.10.021
Notredame, C.-E., Pins, D., Deneve, S., Jardri, R., 2014. What visual illusions teach us about schizophrenia. Front Integr Neurosci 8, 63. https://doi.org/10.3389/fnint.2014.00063
Pantano, P., Baron, J.C., Samson, Y., Bousser, M.G., Derouesne, C., Comar, D., 1986. Crossed cerebellar diaschisis. Further studies. Brain 109 (Pt 4), 677–694. https://doi.org/10.1093/brain/109.4.677
Parenti, A., Jardri, R., Geoffroy, P.A., 2016. How Anti-NMDAR Encephalitis Sheds Light on the Mechanisms Underlying Catatonia: The Neural Excitatory/Inhibitory Imbalance Model. Psychosomatics 57, 336–338. https://doi.org/10.1016/j.psym.2016.01.007
Parr, T., Friston, K.J., 2018. The Anatomy of Inference: Generative Models and Brain Structure. Front Comput Neurosci 12, 90. https://doi.org/10.3389/fncom.2018.00090
Parr, T., Markovic, D., Kiebel, S.J., Friston, K.J., 2019. Neuronal message passing using Mean-field, Bethe, and Marginal approximations. Scientific Reports 9, 1889.

https://doi.org/10.1038/s41598-018-38246-3
Paus, T., Keshavan, M., Giedd, J.N., 2008. Why do many psychiatric disorders emerge during adolescence? Nat Rev Neurosci 9, 947–957. https://doi.org/10.1038/nrn2513
Peer, M., Prüss, H., Ben-Dayan, I., Paul, F., Arzy, S., Finke, C., 2017. Functional connectivity of large-scale brain networks in patients with anti-NMDA receptor encephalitis: an observational study. Lancet Psychiatry 4, 768–774. https://doi.org/10.1016/S2215-0366(17)30330-9
Pouget, A., Beck, J.M., Ma, W.J., Latham, P.E., 2013. Probabilistic brains: knowns and unknowns. Nat Neurosci 16, 1170–1178. https://doi.org/10.1038/nn.3495
Powers, A.R., Mathys, C., Corlett, P.R., 2017. Pavlovian conditioning-induced hallucinations result from overweighting of perceptual priors. Science 357, 596–600. https://doi.org/10.1126/science.aan3458
Price, C.J., Warburton, E.A., Moore, C.J., Frackowiak, R.S., Friston, K.J., 2001. Dynamic diaschisis: anatomically remote and context-sensitive human brain lesions. J Cogn Neurosci 13, 419–429. https://doi.org/10.1162/08989290152001853
Rolls, E.T., Deco, G., 2011. A computational neuroscience approach to schizophrenia and its onset. Neurosci Biobehav Rev 35, 1644–1653. https://doi.org/10.1016/j.neubiorev.2010.09.001
Rolls, E.T., Joliot, M., Tzourio-Mazoyer, N., 2015. Implementation of a new parcellation of the orbitofrontal cortex in the automated anatomical labeling atlas. Neuroimage 122, 1–5. https://doi.org/10.1016/j.neuroimage.2015.07.075
Schwartenbeck, P., FitzGerald, T.H.B., Dolan, R., 2016. Neural signals encoding shifts in beliefs. NeuroImage 125, 578–586. https://doi.org/10.1016/j.neuroimage.2015.10.067
Selemon, L.D., 2013. A role for synaptic plasticity in the adolescent development of executive function. Transl Psychiatry 3, e238. https://doi.org/10.1038/tp.2013.7
Shinn, A.K., Baker, J.T., Cohen, B.M., Öngür, D., 2013. Functional connectivity of left Heschl's gyrus in vulnerability to auditory hallucinations in schizophrenia. Schizophrenia Research 143, 260–268. https://doi.org/10.1016/j.schres.2012.11.037
Sohal, V.S., Rubenstein, J.L.R., 2019. Excitation-inhibition balance as a framework for investigating mechanisms in neuropsychiatric disorders. Mol. Psychiatry. https://doi.org/10.1038/s41380-019-0426-0
Sommer, I.E.C., Diederen, K.M.J., Blom, J.-D., Willems, A., Kushan, L., Slotema, K., Boks, M.P.M., Daalman, K., Hoek, H.W., Neggers, S.F.W., Kahn, R.S., 2008. Auditory verbal hallucinations predominantly activate the right inferior frontal area. Brain 131, 3169–3177. https://doi.org/10.1093/brain/awn251
Stephan, K.E., Friston, K.J., Frith, C.D., 2009. Dysconnection in schizophrenia: from abnormal synaptic plasticity to failures of self-monitoring. Schizophr Bull 35, 509–527. https://doi.org/10.1093/schbul/sbn176
Sterzer, P., Adams, R.A., Fletcher, P., Frith, C., Lawrie, S.M., Muckli, L., Petrovic, P., Uhlhaas, P., Voss, M., Corlett, P.R., 2018. The predictive coding account of psychosis. Biol Psychiatry 84, 634–643. https://doi.org/10.1016/j.biopsych.2018.05.015
Uddin, L.Q., 2020. Bring the Noise: Reconceptualizing Spontaneous Neural Activity. Trends in Cognitive Sciences 0. https://doi.org/10.1016/j.tics.2020.06.003
van den Heuvel, M.P., Sporns, O., Collin, G., Scheewe, T., Mandl, R.C.W., Cahn, W., Goñi, J., Hulshoff Pol, H.E., Kahn, R.S., 2013. Abnormal rich club organization and functional brain dynamics in schizophrenia. JAMA Psychiatry 70, 783–792. https://doi.org/10.1001/jamapsychiatry.2013.1328
Voss, L.J., Baas, C.H., Hansson, L., Steyn-Ross, D.A., Steyn-Ross, M., Sleigh, J.W., 2012. Investigation into the effect of the general anaesthetics etomidate and ketamine on long-range coupling of population activity in the mouse neocortical slice. Eur. J. Pharmacol. 689, 111–117. https://doi.org/10.1016/j.ejphar.2012.06.003
Xiang, Q., Xu, J., Wang, Y., Chen, T., Wang, J., Zhuo, K., Guo, X., Zeljic, K., Li, W., Sun, Y., Wang, Z., Li, Y., Liu, D., 2019. Modular Functional-Metabolic Coupling Alterations of Frontoparietal Network in Schizophrenia Patients. Front. Neurosci. 13. https://doi.org/10.3389/fnins.2019.00040
Yang, G.J., Murray, J.D., Wang, X.-J., Glahn, D.C., Pearlson, G.D., Repovs, G., Krystal, J.H., Anticevic, A., 2016. Functional hierarchy underlies preferential connectivity disturbances in schizophrenia. Proc Natl Acad Sci USA 113, E219-28. https://doi.org/10.1073/pnas.1508436113
Yeh, F.-C., Panesar, S., Fernandes, D., Meola, A., Yoshino, M., Fernandez-Miranda, J.C., Vettel, J.M., Verstynen, T., 2018. Population-averaged atlas of the macroscale human structural connectome and its network topology. NeuroImage 178, 57–68.

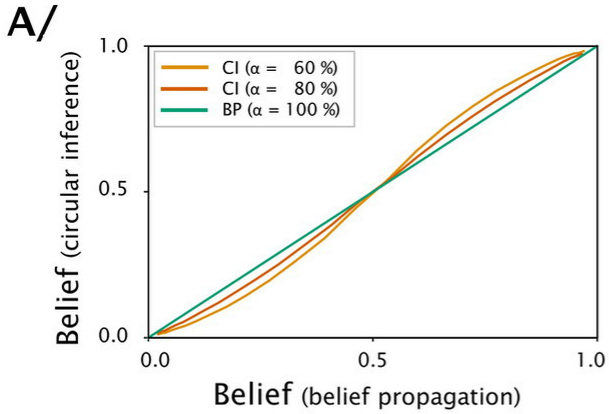
https://doi.org/10.1016/j.neuroimage.2018.05.027
Zalesky, A., Fornito, A., Seal, M.L., Cocchi, L., Westin, C.-F., Bullmore, E.T., Egan, G.F., Pantelis, C., 2011. Disrupted Axonal Fiber Connectivity in Schizophrenia. Biological Psychiatry, N-Methyl-D-Aspartate Receptor Function and Cortical Connectivity in Schizophrenia 69, 80–89. https://doi.org/10.1016/j.biopsych.2010.08.022

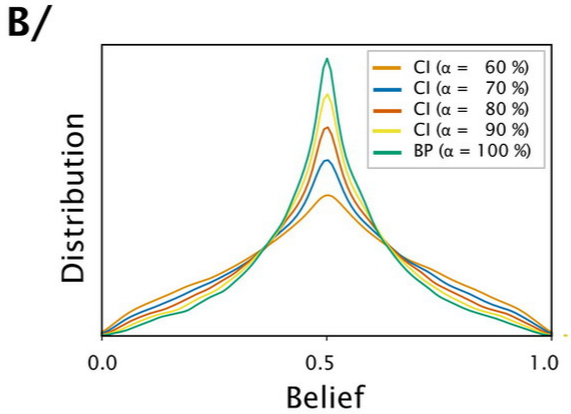

A/ Causal Model
tree
leaf
green
B/ Belief Propagation
Prior
Belief
Sensory input
C/ Circular Inference
i
j
k

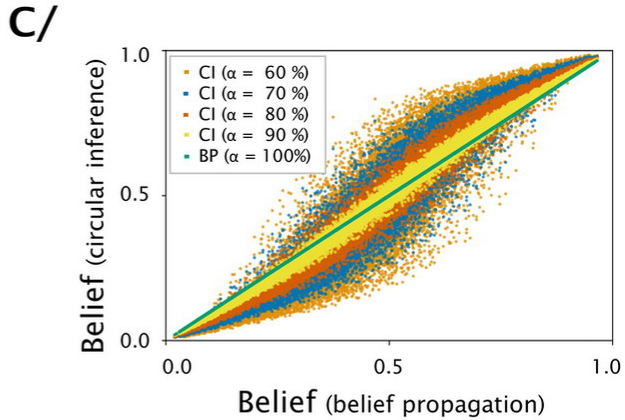

A/
Node Index
Node Index
1
5
10
15
20
25
30
32
B/
t
Module 1
Module 2
Module 3
Module 4
C/
Belief (p(x)=1)
Time (iterations of BP)
0.0
0.5
1.0
0
500
1000
D/
degree of a node
participation coefficient

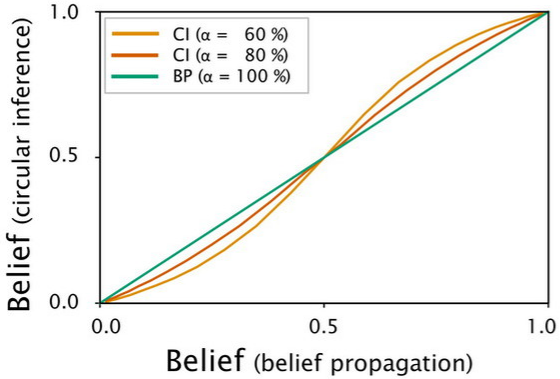
ABSTRACT GRAPHS

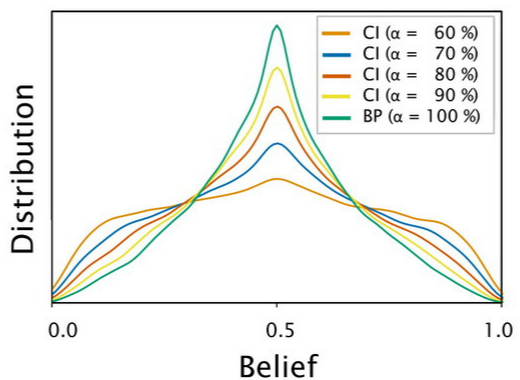
REAL CONNECTOME

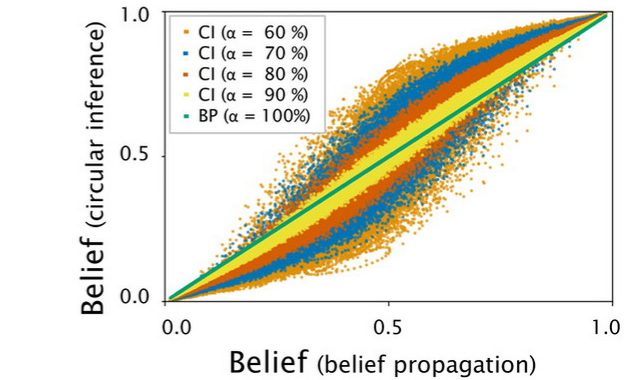




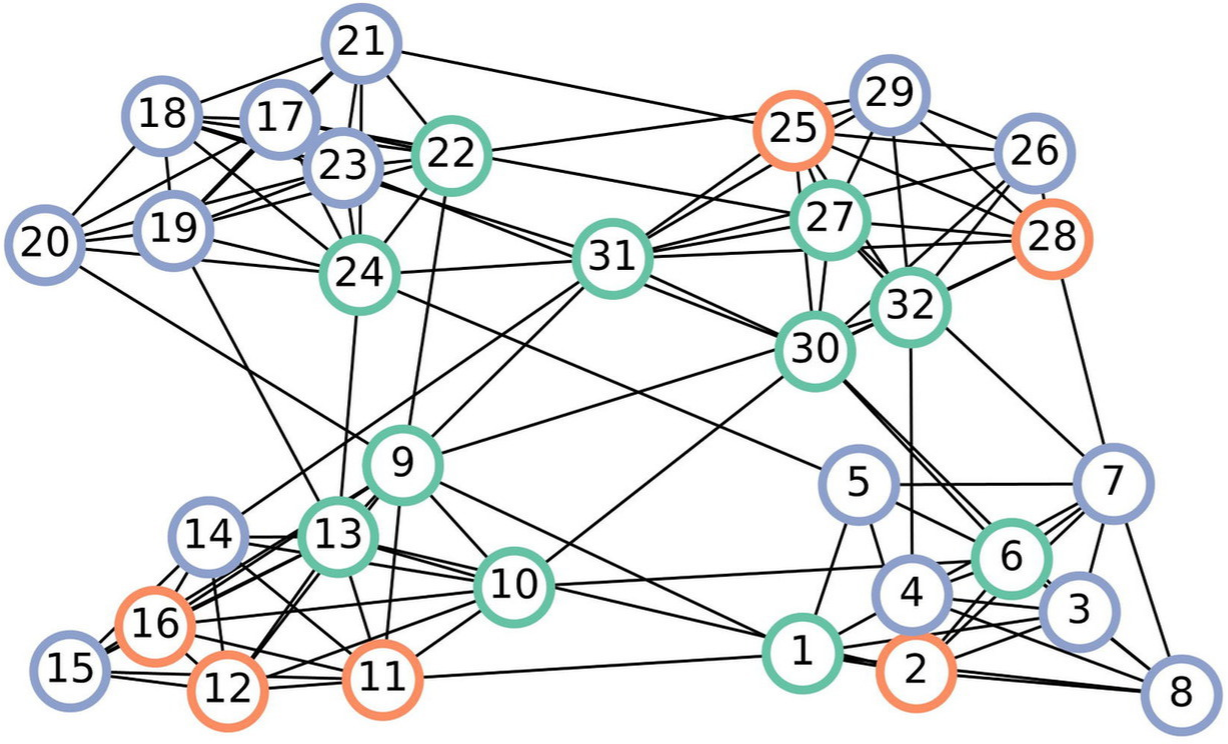


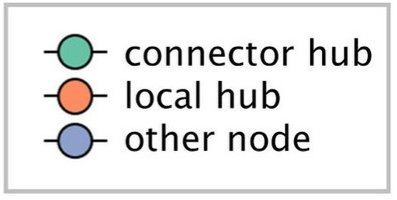




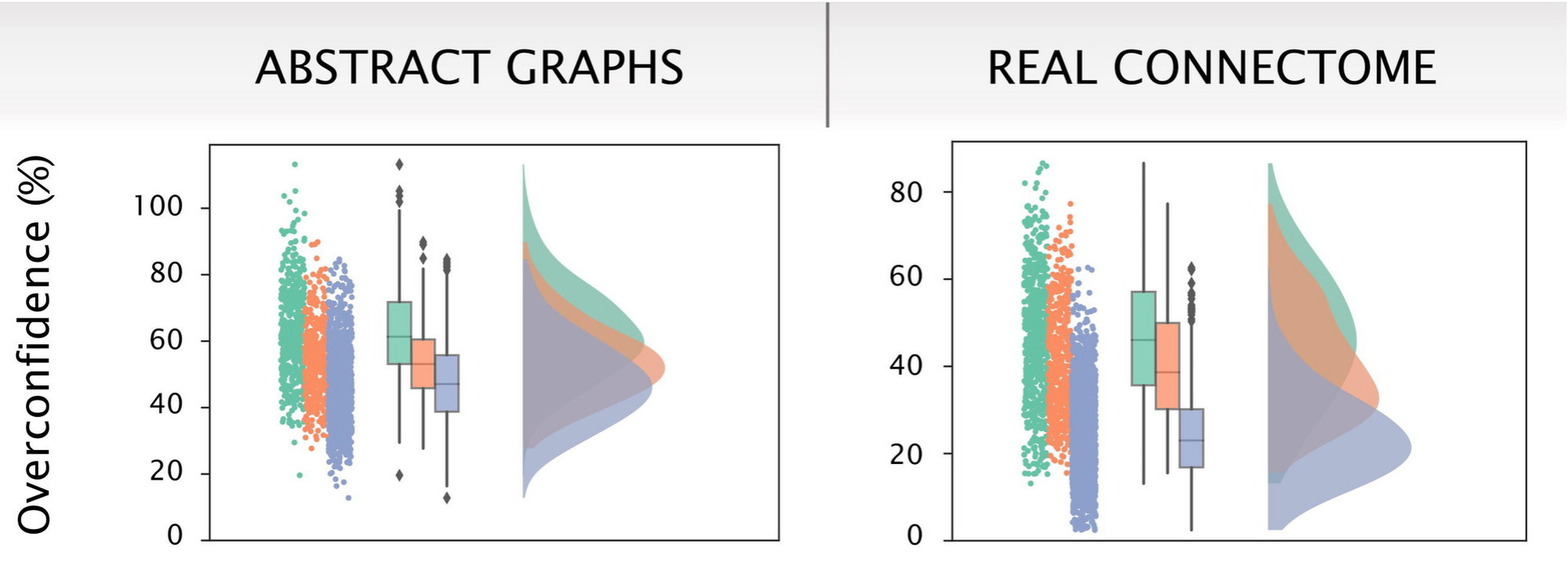


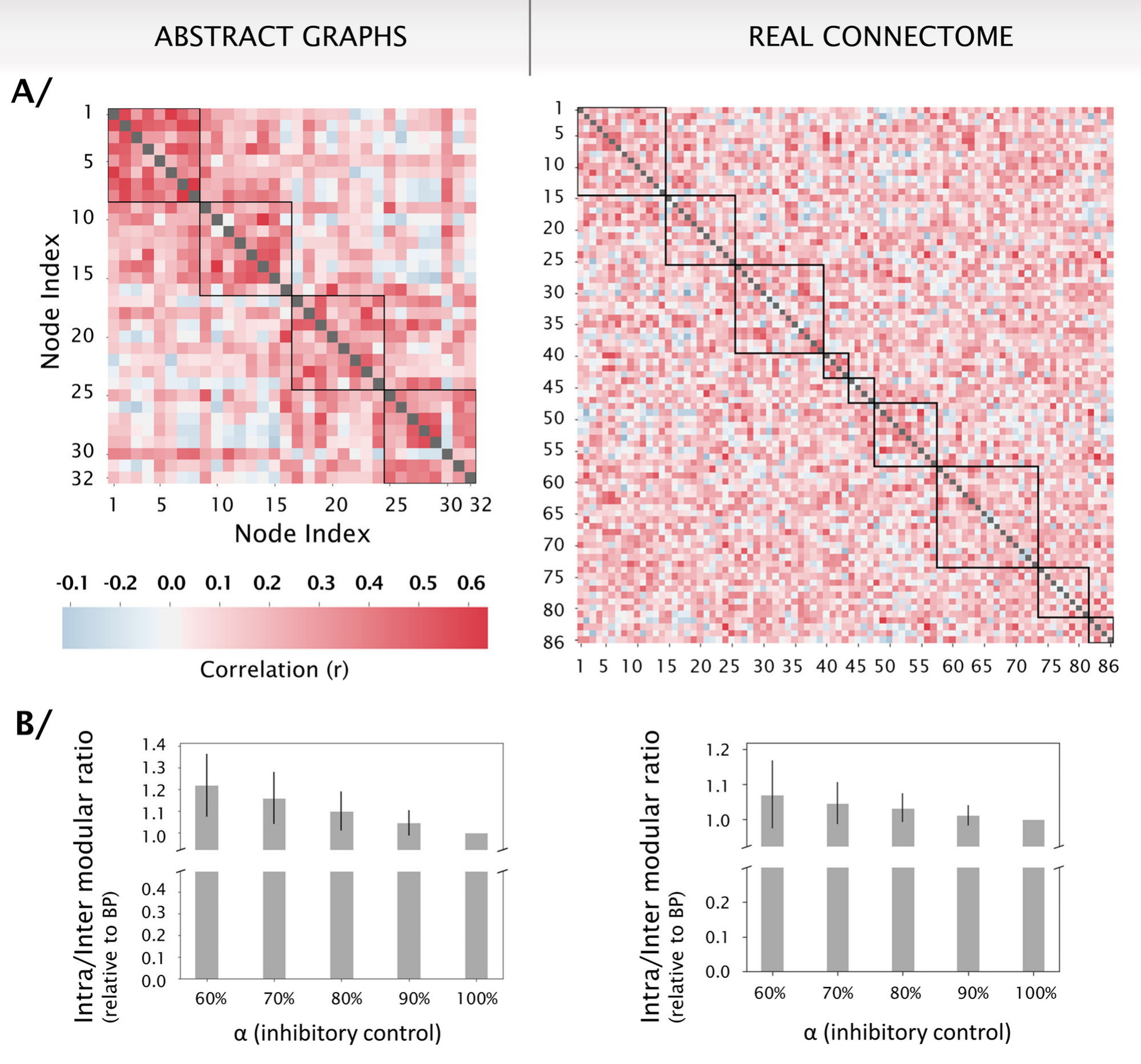
A/


21
18
17
23
22
25
29
26
20
19
24
31
27
28
32
30
9
5
7
14
13
6
10
4
3
16
15
12
11
1
2
8


connector hub
local hub
other node

B/


ABSTRACT GRAPHS
REAL CONNECTOME
Overconfidence (%)
100
80
60
40
20
0
80
60
40
20
0


ABSTRACT GRAPHS
REAL CONNECTOME
A/
Node Index
Node Index
Correlation (r)
B/
Intra/Inter modular ratio
(relative to BP)
α (inhibitory control)
60%
70%
80%
90%
100%